Spring 2018

The build-up of auditory stream segregation in adult cochlear implant users: Effect of differences in frequency and amplitude-modulation rate

Alexandria Matz
James Madison University

The Build-Up of Auditory Stream Segregation in Adult Cochlear Implant Users: Effect of Differences in Frequency and Amplitude-Modulation Rate

Alexandria F. Matz

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial fulfillment of the Requirements

for the degree of

Doctor of Audiology

Communication Sciences and Disorders

May 2018

FACULTY COMMITTEE:

Committee Chair: Yingjiu Nie, Ph.D.

Committee Members/Readers:

Lincoln Gray, Ph.D.

Christopher Clinard, Ph.D.

Acknowledgements

Thank you to my committee for supporting me throughout this journey and making this project possible.

Thank you to my advisor, Dr. Yingjiu Nie, who has been my shoulder to lean on throughout my entire graduate career. Your guidance, kindness, and hard work has made this a seamless and enjoyable process. I couldn't have done it without you.

Special thanks to our wonderful research participants and Harley Wheeler, Lindsey Seyfried, and Sarah Troy for helping out in the research lab.

This work is supported by the Research Grant from the College of Behavioral and Health Studies and supported by the Roger Ruth Memorial Grant from the Department of Communication Sciences and Disorders at James Madison University.

Table of Contents

List of Tables

List of Figures

Abstract

This project will use an objective approach to evaluate the effect of inter-subsequence frequency difference and amplitude-modulation rate on build-up stream segregation in CI users. Six post-lingually deafened CI users, between 18 and 75 years old, have been studied and compared to four normal-hearing listeners, between 18 and 75 years old. Repeated pairs of A and B noise bursts were adopted from a previous work (Nie et al., 2014) with modifications and additional conditions, where A and B bursts are narrow-band noise carrying sinusoidal amplitude modulation (AM). The A and B bursts in a stimulus sequence differed either in the center frequency of the noise band, or in the AM-rate, or both. Subjects identified a deviant in a rhythmic stream and performance (*d'*) reflects the strength of stream segregation. The build-up effect was assessed by comparing performances during long and short sequence durations. Results of this study reveal both CI users and NH listeners showed evidence of build-up effect; however, NH listeners showed stronger stream segregation abilities. Duration has the strongest effect for the 16-10 condition and the weakest effect for the 10-10 condition when both groups were analyzed together. This could indicate that frequency separation is a cue for build-up effect. Frequency separation elicited stream segregation in both CI users and NH listeners. Any amount of frequency separation (within the given conditions) provided cues for stream segregation in NH listeners. Only the largest frequency separation (16-10) provided cues for stream segregation in CI listeners. This could indicate spectral interference still occurs even with 3 channels of separation. Finally, AM-rate separation did not elicit stream segregation in either CI users or NH listeners. These findings are contradictory to previous findings and indicate temporal pitch perception may be used by CI users to separate target auditory streams from background noise.

Chapter 1: Manuscript

Introduction

Auditory stream segregation (also known as auditory streaming) refers to the process that allows listeners to interpret multiple sounds coming from different sources and assign those sounds to individual sound generators (Moore & Gockel, 2012). For example, normal hearing listeners use stream segregation abilities to separate a talker at a noisy party or isolating the violin amongst the other instruments in an orchestra (Nie & Nelson, 2015). Stream segregation depends on the amount of difference between consecutive sounds and the rate of presentation of the sounds (Cooper & Roberts, 2007; Moore & Gockel, 2012; van Noorden, 1975). Researchers have identified two percepts of sound sequences: fusion and fission. As we listen to rapid sequences of sounds, they can be perceived as coming from a single source (fusion or integration) or they can be perceived as coming from two sources (fission or segregation) (Figure 1) (Cooper & Roberts, 2007; Moore & Gockel, 2002).

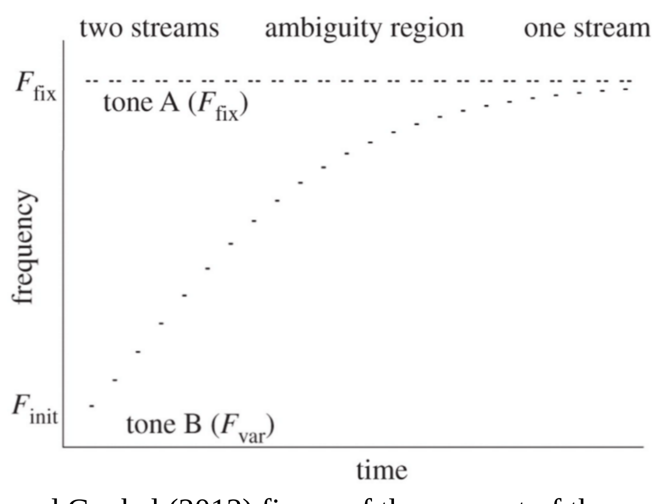

Figure 1. Moore and Gockel (2012) figure of the concept of the ambiguity region.

Cooper and Roberts (2007) describe "tone repetition time" or TRT; the smaller TRT (faster rate), the increased tendency towards stream segregation. Additionally, large frequency separations and high rates of presentation tend to lead toward fission, while small separations and low presentation rates often lead toward fusion (Bockmann-Barthel et al., 2014; Cooper & Roberts, 2007; Moore & Gockel, 2002). The probability of hearing two distinct auditory streams increases as frequency separation between alternating tones increases (Cooper & Roberts, 2007). If the frequency separation is larger than a critical value (the "temporal coherence boundary"), fission always occurs (Figure 2) (Cooper & Roberts, 2007; Moore & Gockel, 2002; van Noorden, 1975).

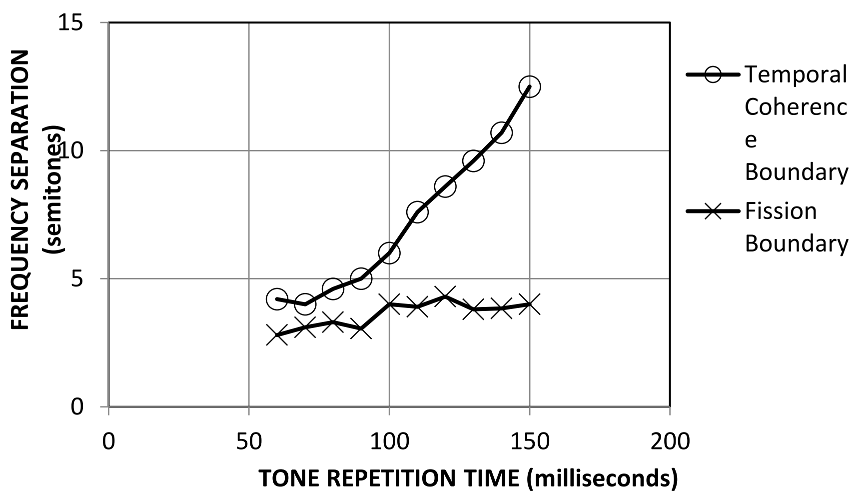

Figure 2. The Fission Boundary (FB) and the Temporal Coherence Boundary (TCB) for auditory stream segregation in van Noorden's unpublished dissertation "Temporal coherence in the perception of tone sequences" (Adapted from van Noorden, 1975).

Listeners will continue to hear two separate streams in this instance, even if instructed to only hear one. This phenomenon has been called "primitive stream segregation" or "obligatory" segregation (Moore & Gockel, 2002). In contrast, if the frequency separation is less than a (different) critical value (the "fission boundary"), then fusion always occurs, even if instructed to listen for two streams (Cooper & Roberts, 2007; Moore & Gockel, 2002; van Noorden, 1975).

Moore and Gockel (2002) examined the ambiguity region (intermediate frequency separations) and how it related to fission and fusion percepts. In their experiment, tone sequences in the form of ABA--ABA… were examined where A and B represented brief sinusoidal tone bursts, and – represented a silent interval. The A and B bursts differed in frequency and were varied to be closer in frequency to one another, or further in frequency from one another. When the sound sequence falls between the fission boundary and the temporal coherence boundary (ambiguity region), listeners report a "flip" spontaneously between integration and segregation (Cooper & Roberts, 2007; Moore & Gockel, 2002). In the ambiguity region, there was a tendency for fission to occur. Fission occurred with increasing exposure time to the sequences. Moore and Gockel (2002) state, "the auditory system starts with the assumption that there is a single sound source, and fission is only perceived when sufficient evidence has built up to contradict this assumption." This build-up effect seems to stabilize after around ten seconds.

Auditory stream formation has been shown to be dependent on the amount of time the target sequence is presented. The tendency for segregation (fission) to occur increases with longer exposure time to the sound sequence. Additionally, the fission percept builds up rapidly over about ten seconds, and builds more slowly up to at least 60 seconds (Moore & Gockel, 2012). As mentioned above, the auditory system assumes that a single sound source is available, but with continuous presentation or build-up, the auditory system can separate streams (fission) (Moore & Gockel, 2012). Attention has been postulated to play a role in the build-up of streaming. Although there are some conflicting ideas, researchers support that build-up of segregation can be decreased by inattention or by a switch of attention (or a combination) (Carolyn et al., 2001; Cusack et al., 2004; Moore & Gockel, 2012).

There have been recent studies that have argued against the presence of the build-up effect. Dieke et al. (2012) challenged the idea that a single sound source is the initial percept for the auditory system. In their experiment, listeners were instructed to indicate their percept as soon as possible and if their percept changed at all during the presented sequence. They revealed that with the largest frequency separations, a two-stream percept was immediately identified by the listeners (probability of 80%), challenging the idea of an initial fusion percept during a sound sequence. Other researchers have reported a similar initial two-stream percept at large frequency separations (Denham et al., 2013). The only increase in two-stream percepts following the initial onset of the sound sequence at the most ambiguous stimulus condition. Dieke et al. (2012) concluded, "a build-up of segregation is not generic and requires ambiguity of the sound sequences to occur" (Bockmann-Barthel et al., 2014).

Congruently, the abilities of cochlear implant users to segregate auditory streams are even less clear. Studies have begun to explore this topic in cochlear implant users, especially when evaluating the build-up effect. Reduced streaming abilities are hypothesized in cochlear implant listeners, as electrical stimulation patterns present similarly to the increased region of excitation along the basilar membrane in hearing-impaired listeners. This wide area of stimulation around each electrode causes an overlap in sound sequences and per the theory of peripheral channeling, fusion is more likely to occur (Marozeau et al., 2013). Cochlear implant users have the ability to understand speech quite well in a quiet environment, but often have a more difficult time in background noise and with music perception. A major limitation of cochlear implant benefit is speech understanding – speech understanding is greatly reduced by competing background sounds. Nelson and Jin (2004) reported competing speech can decrease performance in implant users even at a +16 dB signal-to-noise ratio (Cooper & Roberts, 2007). This may be attributed to the varying spatial interaction between electrode channels (Cooper & Roberts, 2007). The spread of excitation across channels can lead to a high degree of spectral "smearing" and reduced spectral resolution (Cooper & Roberts, 2007).

Several studies have shown evidence toward stream segregation abilities in cochlear implant users. Marozeau et al. (2013) researched the effects of a cochlear implant on different acoustic and perceptual cues responsible for streaming to determine a relationship between these differences in cues and melody segregation. Results overall of this experiment indicated that, "as the physical difference between the target and the distractor increased, listeners on average reported less difficulty segregating the melody from distractor notes" (Marozeau et al., 2013). This finding is relatively consistent with previous research in normal-hearing listeners. This study suggested that like normal-hearing listeners, cochlear implant users can segregate streams when the difference between them is large enough. Other subjective studies of stream segregation abilities in cochlear implant listeners have been conducted. Chatterjee et al. (2006) report similar results; from preliminary studies, CI listeners are able to use intensity differences to segregate stimuli sequences. They used stimuli sequences that were varied (in order to measure build-up) and consisted of two pulse trains which were different in cochlear location. As the frequencies of the sequences became further apart, streaming became stronger (consistent to above results). However, although they found evidence toward streaming abilities in cochlear implant users, no build-up effect was evidenced. Lastly, Bockmann-Barthel et al. (2014) compared cochlear implant users to normal-hearing listeners when evaluating the build-up of streaming. They evaluated the proportion of time for distinguishing the perception of one-stream vs. two-streams and the time needed to make that first perceptual decision. Bockmann-Barthel et al. (2014) concluded similar results between cochlear implant users and normal-hearing listeners – as frequency separations increased, a two-stream percept increased. When evaluating the build-up effect, Bockmann-Barthel et al. (2014) reported the traditional hypothesis of build-up (a two-stream percept emerges from an initial one-stream percept over time), was not supported. That is, users did not default to a one-stream percept at first, then later change their response to a two-stream percept. In fact, in this study, both normal-hearing listeners and cochlear-implant listeners defaulted to favor the two-stream percept initially (Bockmann-Barthel et al., 2014).

Although several studies have provided evidence of stream segregation abilities in cochlear implant users, others have argued that cochlear implant users do not show indication of stream segregation abilities. Cooper and Roberts (2007) reason that cochlear implant listeners must rely on schema-based processing to separate a perceptual stream which puts them at a considerable disadvantage in complex listening environments (especially if attentional resources are limited). This is a contrast to Chatterjee et al. (2006), who suggested that cochlear implant users may have primitive stream segregation abilities.

This current study is a follow-up study to Nie and Nelson (2015). Nie and Nelson (2015) objectively evaluated the role of spectral overlap and amplitude-modulation rate on build-up stream segregation in normal-hearing listeners where the stimuli was constructed to resemble the spectral interaction of signals delivered through a cochlear implant. Previous studies on stream segregation have been subjective; meaning, listeners report the number of auditory streams perceived. Objective measures in relation to stream segregation research involve behavioral responses to stimuli, however the number of streams the listener perceives is not reported. Objective measures involve a listening task in which the listener can select a correct or incorrect response and the strength of stream segregation is reflected by the performance in the listening task. The results of this study suggested that CI users may show stream segregation abilities if spectral separations and amplitude-modulation rate differences are large enough (Nie & Nelson, 2015). Based on these results, the current study was created with similar stimuli constructs (with modifications) and procedures to evaluate the build-up of streaming in cochlear implant users.

Research on these topics continues to be conflicting. This may be due to the differences in methodologies between studies. Most experiments evaluating stream segregation abilities in cochlear implant users have used subjective methods to obtain data. With a lack of these objective measures (described above), data must be evaluated with cautious interpretations. Additionally, differences amongst stimuli constructs and stimuli delivery (direct electrical stimulation vs. soundfield testing) may be reasons for conflicting results. Because cochlear implant users receive degraded spectrotemporal information with auditory inputs due to the limitations of cochlear implant technology, it remains inconclusive whether stream segregation can be elicited in cochlear implant users. This project will use an objective approach (as described above) to evaluate the effect of inter-subsequence frequency difference and amplitude-modulation rate on build-up stream segregation in cochlear implant users.

Materials and Methods

Participants

Ten adult listeners between 18 and 69 years of age, three female, seven male, participated in the study. They were divided into two groups: post-lingually deafened cochlear implant (CI) users (6 participants) age 24 to 69 years with a mean age of 52.5 years, and normal hearing (NH) listeners (4 participants) age 22 to 60 with a mean age of 37.8 years. All NH listeners had symmetric (no greater than a 10 dB difference between ears) audiometric thresholds no greater than 25 dB HL at 250, 500, 1000, 2000, 4000, and 8000 Hz. All CI users wore only one cochlear implant; if they were bilateral users, they were encouraged to wear the CI in the ear they thought they performed best in. If they were bimodal users, they did not use their hearing aid in the other ear. Table 1 (below) illustrates the demographics of each CI user. The Institutional Review Board at James Madison University approved the research procedure to conduct the experiment on human participants. Informed consent was obtained from all participants.

Apparatus

Stimuli were processed through a Lynx 22 soundcard installed in a Dell Optiplex 9010 computer, which ran through a DAC1 device. The analog output of the DAC1 was amplified via a Tucker Davis Technologies, TDT RZ6 system and presented through a Klipsch RB-51 bookshelf speaker. The stimuli were generated using a MATLAB (R2013a) script at a sampling rate of 44,100 Hz. Stimulus presentation and response recording was controlled by the MATLAB script in conjunction with PsychToolbox (version 3) (Brianard, 1997; Pelli, 1997). To record the participants' responses, an RTbox (Li et al., 2010) was used as the hardware interface. Participants were seated at 0^{o} azimuth at a 1-meter distance from the speaker.

Stimulus Sequences

Nine-pair condition (long sequences eliciting build-up): Nine pairs of A and B noise bursts were generated as described in a previous work (Nie & Nelson, 2015) with modifications and additional conditions, where A and B bursts were narrow-band noise carrying sinusoidal amplitude modulation (AM). The A and B bursts in a stimulus sequence differed in either center frequency of the noise band, in the AM-rate, or both. The noise bursts were presented in ABAB sequences, with the duration of an A or B burst being 80 ms including 8 ms rise/fall ramps (with amplitude modulation) with a 50 ms gap between the last offset of a burst to the onset of the next. The full duration of the ABAB sequences was 24.7 seconds. The A bursts (except the initial one) were randomly jittered from their nominal temporal location by a random amount drawn on each presentation ranging from 0 to 40 ms. The center frequencies of A bursts were set at 1803, 3022, or 6665 Hz (for both CI users and NH listeners), the equivalent of the respective center frequencies for a standard Advanced Bionics device's 10th, 13th, and 16th electrodes. The AM-rate alternatives for A bursts were 0, 100, 200, or 300 Hz. The B bursts were presented steadily (with no random temporal jitter) and centered at 1803 Hz, the equivalent of the center frequency of electrode 10 on Advanced Bionics cochlear implant devices. B bursts were presented either at an AM-rate of 50 Hz or without amplitude modulation. The amplitude modulation depth was 100%. Noise bursts were set to the narrowest bandwidths that allow a steady presentation level in the sound field in the participant's location (Walker et al., 1984). Subsequently, bandwidths of 162 Hz were used for the 10th and 13th electrode conditions, and 216 Hz for the 16th electrode conditions. To objectively measure stream segregation abilities based on listeners' behavioral responses, two types of stimulus sequences were adopted. The sequences differed in the placement of the last B burst. In a delayed sequence, the last B burst was delayed from its nominal temporal position by 35 ms. Conversely, in a no-delay sequence, the last B burst was advanced by a randomly drawn amount ranging from 0-10 ms. The total sequence duration was 3.10 seconds for the delayed sequences, and 3.06-3.07 seconds for the no-delay sequences (Nie & Nelson, 2015).

It should be noted that, for CI users with Cochlear and Med EL devices, the center frequencies of A and B bursts were shifted such that they fell in the center of the analysis bands that consisted of the nominal center frequencies. Table 1 shows the specific center frequencies of A and B bursts in different conditions for individual CI users.

Participant Code	Age	CI Model	Center Fq of B bursts corresponding to the 10th channel of an AB device (number of the analysis band with the personal device)	Center Fq of A bursts corresponding to the 13th channel of an AB device (number of the analysis band with the personal device)	Center Fq of A bursts corresponding to the 16th channel of an AB device (number of the analysis band with the personal device)
SD-ACI001	53	Cochlear	1808 (E11)	2927 (E8)	6418 (E2)
EI-ACI005	69	Cochlear	1683 (E12)	2871 (E8)	6485 (E2)
ST-ACI006	69	MedEl	1632 (E7)	3064 (E10)	7352 (E12)
NB-ACI007	43	Cochlear	1741 (E11)	3092 (E7)	6828 (E2)
PX-ACI008	24	Cochlear	1683 (E12)	2871 (E8)	6485 (E2)
SP-ACI009	57	Cochlear	1683 (E12)	2871 (E8)	6485 (E2)

Table 1. Patient demographics and corresponding center frequencies to the reference Advanced Bionics center frequencies.

Four AM-rate conditions for the relationship between the A and B bursts were applied for the NH listeners, while three AM-rate conditions were applied for the CI listeners—the fewer conditions conducted to accommodate for the limited availability of the CI listeners. For the NH listeners, the AM-rates were as follows: unmodulated (AM0-0) with no AM applied to either A or B band; modulation rates one octave apart (AM100-50) with A and B bands modulated at rates of 100 and 50 Hz respectively; modulation rates two octaves apart (AM200-50) with A and B bands modulated at rates of 200 and 50 Hz respectively; and finally, modulation rates 2.585 octaves apart (AM300-50) with A and B bands modulated at rates of 300 and 50 Hz respectively. Amplitude modulation rates for stimuli were established by previous research, which determined elicitation of nonspectral pitch for sinusoidal amplitude modulation (SAM) between frequencies of 40 and 850 Hz (Burns & Viemeister, 1976; Burns & Viemeister, 1981; Fitzgerald & Wright, 2005). Adaptive procedures from Jesteadt (1980) were adopted to account for perceived loudness differences in presentation of frequency-varied stimuli. These procedures approximated the loudness for the A bursts at the 13th (A13) and 16th (A16) electrode equivalents to the loudness for the 10th (A10 and/or B10) electrode equivalent, registered at 60dB A in the soundfield (Wheeler & Nie, 2016). Table 2 displays all stimuli conditions that were examined in this experiment.

Electrode Differences (A-B)	AM-Rate Differences			
16-10	(AM300-50)	(AM200-50)	(AM100-50)	(AM0-0)
13-10	(AM300-50)	(AM200-50)	(AM100-50)	(AM0-0)
10-10	(AM300-50)	(AM200-50)	(AM100-50)	(AM0-0)

Table 2. Displays all possible test conditions through a matrix of parameters. Electrode conditions are displayed to the left (i.e. 16-10 corresponds to A burst centered at electrode 16 and B burst centered at electrode 10). Amplitude-modulation rate differences are displayed to the right (i.e. where (AM300-50) corresponds to the A burst amplitude-modulated at a rate of 300 Hz with the B burst amplitude-modulated at a rate of 50 Hz). The (AM100-50) condition is grayed out as it represents that only NH listeners were tested with that stimulus condition, the other three conditions were applied to both NH and CI groups.

Three-pair condition (short sequences providing baseline for build-up effect): The temporal settings for the A and B bursts in a 3-pair sequence were the same as those in a 9-pair sequence with only the first, second, and the last stimulus pairs of a 9-pair sequence preserved.

Procedure

Each subject performed adaptive loudness balancing (Jesteadt, 1980) to begin testing. Loudness balancing eliminated loudness as a confounding variable between conditions with spectral differences (i.e. A burst at 6665Hz and B burst at 1083 Hz). During this procedure, the participant was seated in the designated location (mentioned above) and two consecutive noise bursts were presented. The participant was then tasked to select either “1” or “2” on a keyboard which corresponded to the perceived louder tone. Based on the participant’s response, the intensity of the target noise burst was adaptively adjusted. When no loudness difference was perceived, the participants were told to take a guess as to which one they perceived as louder. The loudness balancing was continued until noise bursts were matched in loudness. The first set of tasks in the loudness balancing were the 16^{th} electrode (A burst) against the 10^{th} electrode (B burst), where a 16^{th} electrode equivalent burst and 10^{th} electrode equivalent burst were presented, and the participant selected which was louder. This task continued until loudness matching was achieved. Then, this task was repeated with the 13^{th} electrode against the 10^{th} electrode until loudness matching was achieved.

Training session:

The next step included an initial training sequence that would reflect the task of the experimental conditions. Listeners were told to direct attention on segregating the two overlaid streams – this would yield better performance. Performance, d’, was measured through a single-interval, yes/no approach. The listeners were instructed to identify either a delayed sequence or a no-delay sequence after the streams were presented. In order to distinguish the delayed sequence vs. a no-delay sequence, listeners had to discriminate a prolonged gap between the last two B bursts, as opposed to the constant B-to-B gaps of the previous nine B bursts . For each stimulus block, 50% of the trials were randomly selected to consist of delayed sequences while the remaining 50% trials consisted of no-delay sequences (Nie & Nelson, 2015). Training sessions lasted from either 30 minutes to 1 hour, depending on how quickly the participant was able to perform the task (score of d' >1.5 for the Electrode 16-10/AM-rate 0-0 condition).

Participants received the following directions: “You are going to hear a sequence of alternating (A and B) noise bursts. These A and B bursts may differ in some characteristics. Try to focus on separating the two streams, focusing on the lower frequency, steady noise bursts (B bursts). The two streams may sound quite different, or extremely similar, but try to focus on the B stream. As you listen to the “B” stream, follow the stream to the end, and determine if the last burst was either delayed from the previous bursts, or not delayed from the previous bursts. As soon as you determine the delayed vs. no-delay sequence, select “1” or “4” on the response (RTbox) box.”

The number “1” on the RTbox corresponded with a delayed perception, while the “4” on the RTbox corresponded with a no-delay perception. A computer screen on the patient’s side displayed a red color over the delayed or no-delay window if the participant chose incorrectly, and a green color over the delayed or no-delay window if the participant chose correctly. This measure was implemented to inform subjects of their performance. The delay between the last B burst and the previous bursts was changed to 50 ms to start, to make the delay extremely obvious. As performance improved, the amount of possible delay was decreased until performance stabilized at 35 ms (the amount of delay for experimental conditions). A d’ of >1.5 during training sessions indicated the participant correctly understood the task. If a d’ of >1.5 could not be obtained during initial training, training continued, with additional instructions, to focus the participant on the goal of the task. During training, frequency differences of 16 (A electrode) vs. 10 (B electrode) were used. This is because the greater frequency difference was hypothesized to make it easier for the subject to separate streams. The AM-rate used during training conditions was (AM0-0). Both nine-pair and three-pair sequences were trained. Once the participant displayed training was effective, experimental conditions were started.

Experimental conditions:

Participants were instructed to perform the experimental conditions like the training sessions. Data collection lasted from 3.5 hours to 4.5 hours depending on the number of breaks the participant needed in between experimental blocks. Except, the possible delay of the last B burst was fixed at 35 ms. This value was determined based off performance while collecting pilot data. Experimental conditions included two 40-trial blocks of each nine-pair and three-pair conditions. This experiment evaluated the effect of intersubsequence frequency differences and amplitude-modulation rate as cues for the build-up effect. Table 2 above illustrates all the possible experimental conditions. A total of three duration/spectral conditions were assessed in a randomized order for each participant. All four of the AM-rate separations (three for CI users) were nested under each of these four spectral conditions (Nie & Nelson, 2015).

The jittered timing of the A bursts introduced uncertainty to detecting the delayed vs. no-delay ending B burst in a sequence. Because of this, participants could not simply rely on the A-to-B gap as a cue for identifying the delayed vs. no-delay ending B burst. Participants were thus required to follow the B burst sequence and ignore the A burst sequence in order to determine gaps between B bursts (Nie & Nelson, 2015). Stream segregation was objectively measured by the participant’s performance detecting the delayed vs. no-delay B burst. In other words, “for better performance, listeners presumably made mental efforts to segregate B bursts from A bursts to form a perceptual stream of B bursts” (Nie & Nelson, 2015).

When the ceiling performance was reached (i.e., 100% for hit rate and 0% for false alarm rate) in a given block, the d' value was derived following the Equations 1 and 2 (Macmillan & Creelman, 2005) to correct for the ceiling effect where S and N represent the total possible numbers of trials presented for signal and reference sequences, respectively.

$$\text{Hit Rate} = (S-1)/S \times 100\% \quad (1)$$

$$\text{False Alarm Rate} = 1/N \times 100\% \quad (2)$$

Data Analysis

IBM SPSS statistics version 23 was used for data analysis and means and standard errors are reported in the results section. Data were analyzed using a general linear model univariate approach specified in the results section for readability.

Results

Cochlear implant listeners

A 3-way repeated measures (RM) ANOVA was used to examine the within-subject effect of frequency separation, AM-rate, and duration on the sensitivity of detecting the delayed sequences (d'). In the cochlear implant listeners, there was a significant effect of frequency separation, [$F(1.4,14.9) = 53.42$, $p < 0.001$] (Figure 3).

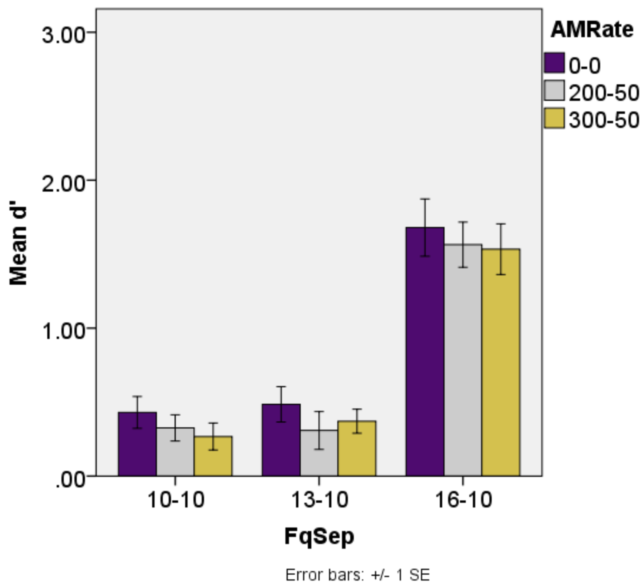

Figure 3. Frequency separation conditions are indicated on the x-axis, mean *d'* scores are indicated on the y-axis; AM-rate conditions are indicated by the colored bars, where purple represents the 0-0 AM-rate condition (no modulation), gray represents the 200-50 AM-rate condition, and gold represents the 300-50 AM-rate condition. Performance (*d*') increases in the 16-10 frequency separation condition. Standard error bars are +/- 1 SE.

Pairwise comparisons with Bonferroni adjustment indicated 16-10>13-10=10-10, $p < 0.001$. This represents a greater effect of frequency separation for the 16-10 sequences than both the 13-10 ($p < 0.001$) and 10-10 sequences ($p = 0.001$). However, performance for the 13-10 and 10-10 was not significantly different, $p = 1.00$. There was no significant effect of AM-rate F(2,22)= 2.99 p=0.071. There was a significant effect of duration, F(1,11)=28.24, p <0.001, (Figure 4) which indicated significantly better performance during the 9-pair sequences than the 3-pair sequences.

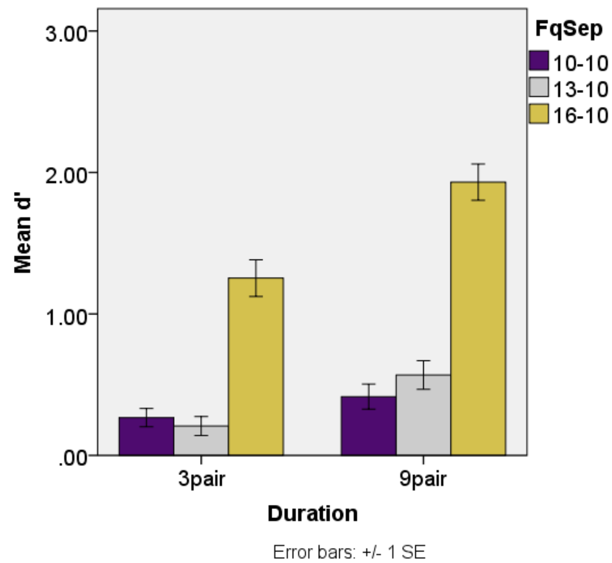

Figure 4. Duration conditions are indicated on the x-axis (3pair represent the short sequences, 9pair represent the long sequences), mean *d'* scores are indicated on the y-axis; frequency separation conditions are indicated by the colored bars, where purple represents the 10-10 (no frequency separation) condition, gray represents the 13-10 condition, and gold represents the 16-10 condition. Performance (*d*') increases in the long duration sequences. Standard error bars are +/- 1 SE.

In addition, Figure 4 shows there was a significant interaction between frequency separation and duration, F(2,22)=4.22, p=0.028, as well as significant interaction between AM-rate and duration, F(1.6,17.5) = 4.39, p = 0.036 (Figure 5).There was no significant interaction between frequency separation and AM-rate (F(4,44)=0.13, p=0.97) or significant three-way interaction between frequency separation, AM-rate, and duration, F(4,44)=0.576, p=0.681.

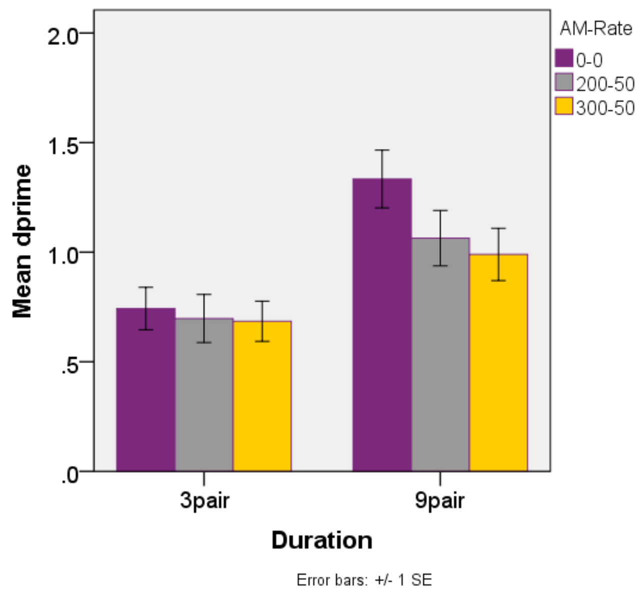

Figure 5. Duration conditions are indicated on the x-axis (3pair represent the short sequences, 9pair represent the long sequences), mean *d'* scores are indicated on the y-axis; AM-rate separation conditions are indicated by the colored bars, where purple represents the 0-0 (no AM-rate separation) condition, gray represents the 200-50 condition, and gold represents the 300-50 condition. Performance (*d*') increases in the long duration sequences, more so for the 0-0 condition than for the other two conditions. Standard error bars are +/- 1 SE.

Normal hearing listeners

A mixed model analysis was used to examine the effect of three fixed factors--frequency separation, AM-rate, and duration on the sensitivity of detecting the delayed sequences (d'). Individual subjects were analyzed as the random factor. In the normal hearing listeners, there was a significant effect of frequency separation, $F(2,164) = 82.31$, $p < 0.001$ (Figure 6).

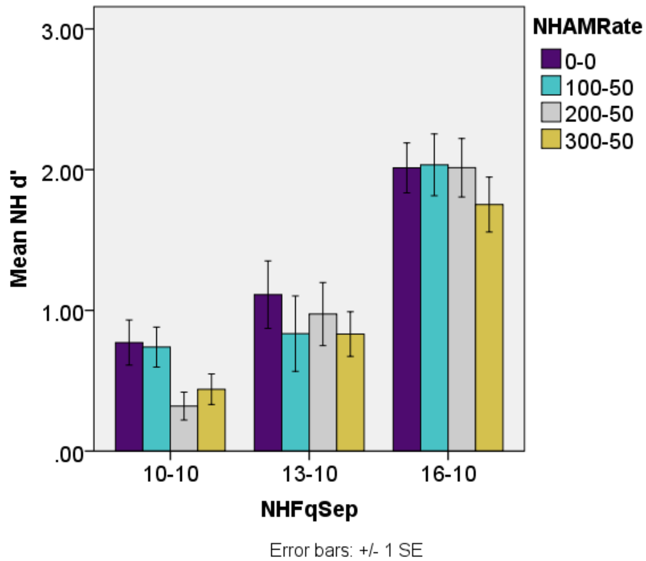

Figure 6. Frequency separation conditions are indicated on the x-axis, mean *d'* scores are indicated on the y-axis; AM-rate conditions are indicated by the colored bars, where purple represents the 0-0 AM-rate condition (no modulation), teal represents the 100-50 AM-rate condition (which was examined in the NH listener group only), gray represents the 200-50 AM-rate condition, and gold represents the 300-50 AM-rate condition. Performance (*d*') increases as frequency separation increases. Standard error bars are +/- 1 SE.

Unlike the cochlear implant listeners, pairwise comparisons indicated 16-10>13-10>10-10. This indicated that any amount of frequency separation improved performance ($p < 0.05$). There was no significant effect of AM-rate, $F(3,164)= 1.91$, $p=0.129$. There was a significant effect of duration, $F(1,164)=32.47$, $p < 0.001$ (Figure 7). Again, like the cochlear implant listeners, pairwise comparisons with Bonferroni adjustment indicated significantly better performance during the 9-pair sequences than the 3-pair sequences.

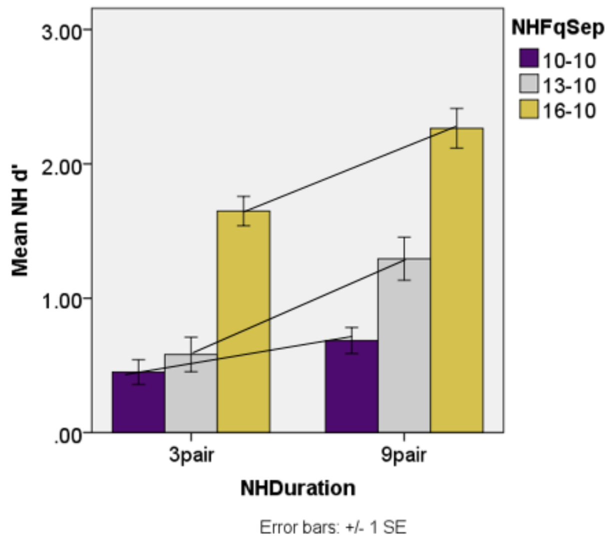

Figure 7. Duration conditions are indicated on the x-axis (3pair represent the short sequences, 9pair represent the long sequences), mean *d'* scores are indicated on the y-axis; frequency separation conditions are indicated by the colored bars, where purple represents the 10-10 (no frequency separation) condition, gray represents the 13-10 condition, and gold represents the 16-10 condition. Performance (*d*') increases in the long duration sequences. Connecting lines show no significant interaction between frequency separation and duration (parallel). Standard error bars are +/- 1 SE.

Unlike the cochlear implant listeners, Figure 6 shows there was no significant interaction between frequency separation and duration, $F(2,164)=2.56$, $p=0.08$. There was also no significant interaction between AM-rate and duration ($F(3,164)=1.115$, $p=0.344$) or frequency separation and AM-rate ($F(6,167)=0.771$, $p=0.594$); there was also no significant three-way interaction between frequency separation, AM-rate, and duration, $F(6,164)=1.103$, $p=0.363$.

Comparison between groups

A 4-way RM-ANOVA was used to examine the effect of—the between-subject factor of listener group, and the within-subject factors of frequency separation, AM-rate, and duration—on the sensitivity of detecting the delayed sequences (d’). When comparing between groups, CI listeners showed significantly lower d’ values than normal hearing listeners, $F(1, 324)=27.99$, $p<0.001$ (Figure 8). Figure 9 also shows a significant interaction between group and frequency separation, $F(2,324)=3.11$, $p<0.05$.

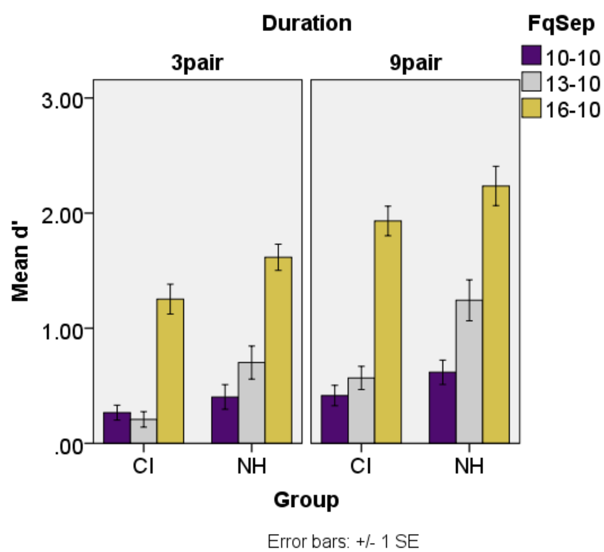

Figure 8. Groups are indicated on the x-axis (CI and NH), mean *d’* scores are indicated on the y-axis; frequency separation conditions are indicated by the colored bars, where purple represents the 10-10 (no frequency separation) condition, gray represents the 13-10 condition, and gold represents the 16-10 condition. The figure is also separated into the duration conditions (3pair and 9pair) for easier comparison. Performance (*d*’) is better for the NH listeners across all conditions the for the CI listeners. Standard error bars are +/- 1 SE.

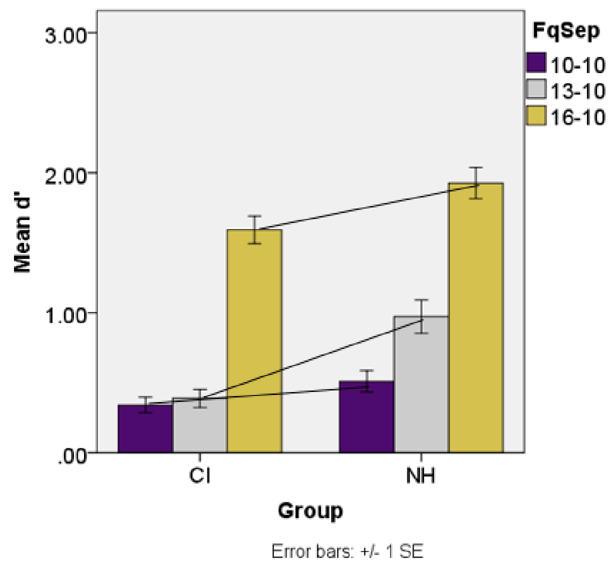

Figure 9. Groups are indicated on the x-axis (CI and NH), mean *d'* scores are indicated on the y-axis; frequency separation conditions are indicated by the colored bars, where purple represents the 10-10 (no frequency separation) condition, gray represents the 13-10 condition, and gold represents the 16-10 condition. The connecting lines show the significant interactions between group and frequency separation. Standard error bars are +/- 1 SE.

When data were collapsed across listener groups (CI and NH listeners), there was a significant effect of frequency separation, $F(2,36)=58.32$, $p<0.001$ and duration, $F(1,18)=42.12$, $p<0.001$ (Figure 10). Post-hoc tests revealed any amount of frequency separation (16-10>13-10>10-10) improved performance (when both groups were analyzed together) significantly ($p<0.05$). Additionally, Figure 10 shows a significant interaction between frequency separation and duration, $F(2, 36)=3.40$, $p< 0.05$.

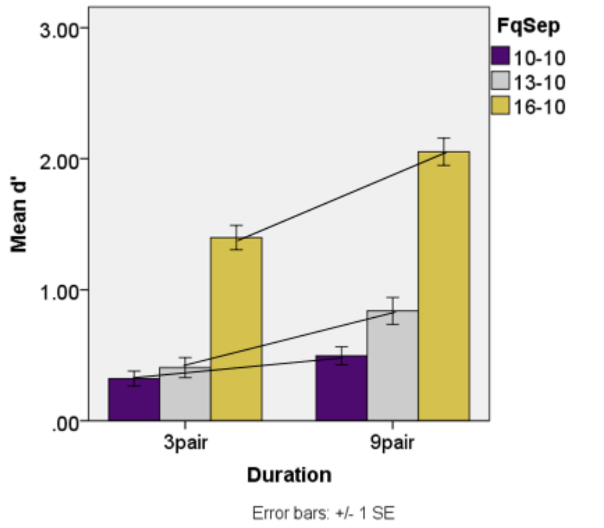

Figure 10. Duration conditions are indicated on the x-axis (3pair represent the short sequences, 9pair represent the long sequences), mean *d'* scores are indicated on the y-axis; frequency separation conditions are indicated by the colored bars, where purple represents the 10-10 (no frequency separation) condition, gray represents the 13-10 condition, and gold represents the 16-10 condition. Performance (*d*') increases in the long duration sequences. The connecting lines represent the significant interaction between frequency separation and duration. Standard error bars are +/- 1 SE.

There was also a significant effect of AM-rate, F(2,36)=8.74, p =0.001 (Figure 11). Post-hoc tests also revealed that the both AM-rate differences (AM300-50 and AM200-50) significantly interfered with performance (p<0.05) when compared to no AM-rate applied to the stimuli (AM0-0).

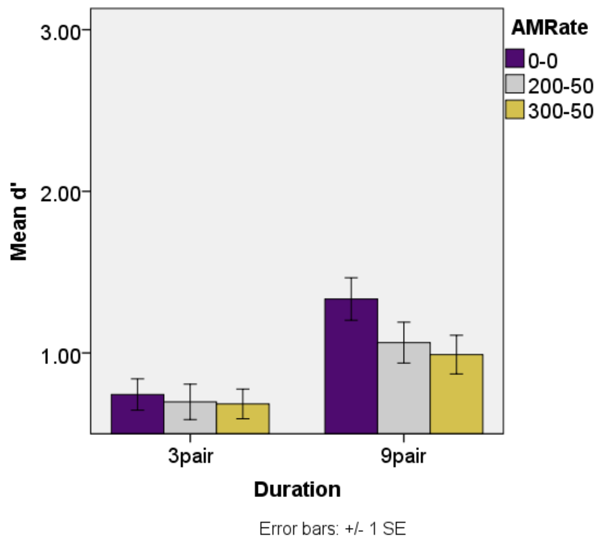

Figure 11. Duration conditions are indicated on the x-axis (3pair represent the short sequences, 9pair represent the long sequences), mean *d'* scores are indicated on the y-axis; AM-rate conditions are indicated by the colored bars, where purple represents the 0-0 (no modulation) condition, gray represents the 200-50 condition, and gold represents the 300-50 condition. Performance (*d*') was best in the 0-0 AM-rate condition and decreased with increasing AM-rate. Standard error bars are +/- 1 SE.

There was no significant difference between AM300-50 and AM200-50 (p=1.000). Finally, there was no significant four-way interaction (group x frequency separation x AM-rate x duration), F(4,72)=0.279, p=0.891.

Discussion

In the current study, we used an objective stimulus paradigm to evaluate whether cochlear implant listeners could segregate two interleaved auditory subsequences of noise bursts in two auditory streams based on differences in frequency, amplitude-modulation rate, or in both. The results of this study are consistent with hypotheses from Nie and Nelson (2015) that spectral cues can elicit stream segregation in cochlear implant users. Our results indicated some differences between studies and will be explained further below. Additionally, our results extend previous findings that cochlear implant users do show evidence for the build-up of stream segregation when spectral differences are large enough.

Stream Segregation Based on Spectral Separations and AM-Rate Differences in Cochlear Implant Users

Consistent with Bockmann-Barthel et al., (2014), who supported large frequency separations tend to lead toward fission through their subjective experiments, our findings showed that large frequency separations (~1700 to ~6500 Hz) could elicit stream segregation using an objective paradigm. Our study also minimized any intensity cues as a means for stream segregation as loudness balancing between A and B bursts was performed, which had been shown to influence streaming abilities in Bregman et al., (1999). In cochlear implant users, only the largest amount of frequency separation improved performance. These results may be consistent with the theory of "peripheral channeling," as the closer in frequency the noise bursts were, the more likely current spread across electrodes occurred causing spectral smearing, which reduced streaming abilities.

In contrast to Nie and Nelson (2015), who found that AM-rate differences aided in stream segregation, our study revealed that there was no significant effect of AM-rate on performance in cochlear implant users. That is, no matter the amount of AM-rate differences, cochlear implant users did not use these temporal cues as a means for stream segregation. From this, we concluded that spectral separations are more important for cochlear implant users to separate streams than changes in AM-rate separations.

Stream Segregation Based on Spectral Separations and AM-Rate Differences in Normal-Hearing Listeners

To objectively evaluate stream segregation abilities in cochlear implant listeners, the current study used normal-hearing listeners to compare results. The normal-hearing listeners underwent the same experiment with one additional parameter (AM-rate 100-50). Like cochlear implant listeners, results indicated a significant effect of frequency separation in normal-hearing listeners. But, in contrast to cochlear implant listeners, any amount of frequency separation improved performance, although the largest frequency separation had the best performance. Consistent with Moore and Gockel (2002), as frequency separations increased, the probability of hearing two streams increased. These results may also support the “peripheral channeling” theory as normal hearing listeners experienced less spectral overlap in excitation patterns than cochlear implant listeners with a same amount of separation between the center frequency of noise bands. Additionally, this result is consistent with the conclusions by Bregman et al. (2001) who stated that overlap of excitation patterns does not necessarily prevent segregation.

Interestingly, in contrast to Nie and Nelson (2015), who found that AM-rate differences aided in stream segregation, our study revealed that there was no significant effect of AM-rate on performance in normal hearing listners. That is, no matter the amount of AM-rate differences, normal-hearing listeners did not use these temporal cues as a means for stream segregation. In Nie and Nelson (2015), wideband noise carriers were used, whereas the current study used narrowband noise carrier. Lemańska, Sęk, & Skrodzka (2002) reported a significant lower threshold in amplitude-modulation rate discrimination with wideband noise carriers than with narrowband noise carriers. They proposed that the larger intrinsic amplitude fluctuations associated with the narrowband noise may have interfered with the amplitude changes resulted from the amplitude modulation.

Stream Segregation Based on Spectral Separations and AM-Rate Differences Between Groups

When comparing the cochlear implant listeners to the normal-hearing listeners, the cochlear implant listeners showed significantly lower d’ values than the normal hearing listeners. This indicated that overall, normal-hearing listeners had better streaming abilities than cochlear implant listeners. Results also indicated a significant interaction between group and frequency separation. That is, frequency separation had a stronger effect for the normal-hearing listeners than the cochlear implant listeners. This interaction is consistent with the results above as any amount of frequency separation increased performance in normal-hearing listeners compared to only the largest frequency separation increasing performance for the cochlear implant listeners. Additionally, we also analyzed the data collapsed amongst all participants (CI and NH listeners together). When data were collapsed across listener groups, there was an effect of frequency separation. These results showed that any amount of frequency separation improved performance. In contrast to previous results, there was a significant effect of amplitude-modulation rate. This could be attributed to the total *n* used when data from groups were combined. These results suggest that with a larger sample size, amplitude-modulation rate effects could be significant. However, the significant effect of amplitude-modulation rate was not the same as found by Nie and Nelson (2015). In the current study, amplitude-modulation rate differences significantly interfered with performance. That is, as amplitude-modulation rate differences increased, performance decreased. Interestingly, increased amplitude-modulation rate differences caused stream segregation abilities to decrease.

Build-up Stream Segregation in Cochlear Implant Users compared to Normal-Hearing Listeners

Both cochlear implant listeners and normal hearing listeners showed evidence of build-up stream segregation as performance improved with the 9-pair condition relative to the 3-pair condition. Nie and Nelson (2015) used 12-pair sequences to elicit build-up, whereas the current study used 9-pair sequences. This study showed that even 9-pair sequences were long enough to elicit build-up stream segregation when compared to 3-pair sequences. However, cochlear implant listeners showed a significant interaction between frequency separation and duration, where normal hearing listeners did not. Deike et al. (2012) argued that when frequency differences between the two stimulus sequences are adequately robust, listeners may form auditory streams instantaneously with the onset of the stimulus, thus do not need to take time to build up the stream segregation. The lack of interaction of frequency separation and duration in normal hearing listeners, together with the monotonically increased d’ as a function of increased frequency-separation, is consistent with the notion in Deike et al. (2012). In other words, the amount of frequency-separation selected in the current study was adequately robust for the normal-hearing listeners to segregate the streams of A and B instantaneously. But, because of the significant interaction between frequency separation and duration, cochlear implant users show evidence that the build-up of stream segregation is facilitated by increased frequency separations. This interaction indicates that as the frequency separation increased, listeners made more improvement with lengthening the duration of stimulus sequences. Thus, the interaction confirms that the frequency separation contributed to the build-up effect, in turn, was a cue for stream segregation. These results contrast studies who have reported no evidence of build-up in cochlear implant users (Chatterjee et al., 2006; Cooper and Roberts, 2009). The discrepancy between studies will be discussed further in this section.

The current study also examined the build-up effect when the groups were collapsed together. Evidence of build-up stream segregation is again supported as performance increased in the 9-pair sequences when compared to the 3-pair sequences. There was also a significant interaction between frequency separation and duration. This indicates that any amount of frequency separation was a cue for stream segregation with increased performance in the long duration sequences compared to the short duration sequences.

In summary, both cochlear implant users and normal-hearing listeners showed evidence of build-up effect; however, normal-hearing listeners showed stronger stream segregation abilities. Additionally, duration has the strongest effect for the 16-10 condition and the weakest effect for the 10-10 condition when both groups were analyzed together. This could indicate that frequency separation is a cue for build-up effect. Also, frequency separation elicited stream segregation in both cochlear-implant users and normal-hearing listeners. Although, any amount of frequency separation (within the given conditions) provided cues for stream segregation in normal-hearing listeners. Only the largest frequency separation (16-10) provided cues for stream segregation in cochlear-implant users. This could indicate frequency interaction still occurs even with 3 channels of separation, supporting the peripheral channeling theory. Finally, amplitude-modulation rate separation did not elicit stream segregation in either cochlear implant users or normal-hearing listeners. Although, there was an interaction between duration and AM-rate when the CI participants were analyzed together; AM-rate differences hindered the formation of stream segregation.

Possible Alternative Explanations of Results

Because of the variation in methodologies used amongst studies, consistent results have not been found. The manner in which data is collected, has had an influence on results; that is, using subjective vs. objective measures to assess build-up of stream segregation. Additionally, how the stimulus sequences are constructed plays a role in obtained results. This study used an objective paradigm to provide clear evidence for the build-up of stream segregation in cochlear implant users.

Nie and Nelson (2015) offer several explanations as to other cues that may have been involved in stream segregation. First, rhythmic cues are present in the stimulus sequences and could facilitate stream segregation. Rhythmic cues could be used by the listeners to segregate a steadily presented B stream from a temporally jittered A stream. Bendixen (2014), Devergie et al. (2010), and Nie et al., (2014) agree that rhythmic cues have been reported to aid voluntary stream segregation in both neurophysiological and behavioral studies. However, this rhythm-based segregation cannot explain the increased d’ values for the large frequency separations. This improved performance indicated that listeners segregated the A and B streams based on spectral differences.

Next, listeners may have been able to detect the delayed (or no-delay) sequence by focusing on the last pair of A and B bursts (instead of following the entire sequence). Nie and Nelson examined this hypothesis with an ideal observer and data indicated “limited or no reliance on simple gap detection for the last pair of A-B bursts, and supports the stream segregation hypothesis” (Nie & Nelson, 2015).

Amplitude-modulation rate was also shown (when data were collapsed across groups) to hinder stream segregation abilities. Although in addition to the AM-rate difference, the amplitude modulation may have introduced spectral cues by generating distortion products, these cues are more likely to make A and B bursts more different perceptually, in turn, promoting stream segregation.

The bandwidths of the stimulus bursts with the amplitude modulation were computed to examine the spectral spread for the normal-hearing listeners. With center frequencies of 1803, 3022, and 6665 Hz for the electrodes 10, 13, and 16 conditions, the corresponding critical auditory bands were 1583-2022 Hz, 2671-3372 Hz, and 5920-7409 Hz, derived using Equation 3.

$$ERB=24.7(4.37F+1), \quad \text{Equation 3}$$

Where the ERB represents equivalent rectangular band and the value is specified in Hz, but F is in kHz. Due to the signal processing that, following the superimposition of amplitude modulation, the A or B bursts were filtered using a bandpass filter to limit the bandwidth to two times of the modulation rate, only amplitude-modulation rate of 300 Hz would “splatter” the spectrum of some bursts beyond an auditory critical band. Specifically, derived from the resulting noise passbands calculated for the condition with the modulation rate of 300 Hz, the cutoff frequencies for the three electrode conditions are respectively 1503-2103 Hz, 2722-3322 Hz, and 6365-6965 Hz. Thus, only with the E10 stimulus burst, the AM at 300 Hz may lead to the resulting spectrum spreading outside of its corresponding critical auditory band between 1583-2022 Hz. However, this spectral spread would lead to the E10 bursts with AM at 300 Hz (i.e., A bursts) to be more perceptually different than the E10 bursts with AM at 50 Hz (i.e., B bursts), the latter of which has a spectrum confined within the critical auditory band. On the other hand, with the other two stimulus bands (i.e., E13 and E16 conditions), the spectral distortions generated by the AM do not spread outside of the respective critical bands. Thus, the effect of AM-rate differences on stream segregation for these conditions was not likely to be confounded by the spectral spread generated by the amplitude modulation. In consequence, a follow-up study should be conducted evaluating other explanations on the interference of AM-separation on stream segregation.

APPENDIX

Extended Review of Literature

Overview of auditory stream formation

Auditory stream segregation (also known as auditory streaming) refers to the process that allows listeners to interpret multiple sounds coming from different sources and assign those sounds to individual sound generators (Moore & Gockel, 2012). For example, normal hearing listeners use stream segregation abilities to separate a talker at a noisy party or isolating the violin amongst the other instruments in an orchestra (Nie & Nelson, 2015). Stream segregation depends on the amount of difference between consecutive sounds and the rate of presentation of the sounds (Cooper & Roberts, 2007; Moore & Gockel, 2012; van Noorden, 1975). By alternating elements (e.g., tones, noise bursts, phonemes, etc.) of individual streams in the stimuli, researchers have identified two percepts of sound sequences: fusion and fission. As we listen to rapid sequences of sounds, they can be perceived as coming from a single source (fusion or integration) or they can be perceived as coming from two sources (fission or segregation) (Cooper & Roberts, 2007; Moore & Gockel, 2002). Cooper and Roberts (2007) describe "tone repetition time" or TRT (see also in van Noorden, 1975); the smaller TRT (faster rate), the increased tendency towards stream segregation. Moore and Gockel (2002) examined the role of cochlear filtering and excitation patterns in fission and fusion percepts. Moore and Gockel (2002) mentioned several factors influencing sequential stream segregation. First, frequency was determined to play a role in stream segregation. Large frequency separations and high rates of presentation tend to lead toward fission, while small separations and low presentation rates often lead toward fusion (Bockmann-Barthel et al., 2014; Cooper & Roberts, 2007; Moore & Gockel, 2002). The probability of hearing two distinct auditory streams increases as frequency separation between alternating tones increases (Cooper & Roberts, 2007). If the frequency separation is larger than a critical value (the "temporal coherence boundary"), fission always occurs (Cooper & Roberts, 2007; Moore & Gockel, 2002; van Noorden, 1975). Listeners will continue to hear two separate streams in this instance, even if instructed to only hear one. This phenomenon has been called "primitive stream segregation" or "obligatory" segregation (Moore & Gockel, 2002). In contrast, if the frequency separation is less than a (different) critical value (the "fission boundary"), then fusion always occurs, even if instructed to listen for two streams (Cooper & Roberts, 2007; Moore & Gockel, 2002; van Noorden, 1975).

But what happens to the fission and fusion percepts with intermediate frequency separations? This ambiguity region was assessed by (Moore & Gockel, 2002). In their experiment, tone sequences in the form of ABA--ABA… were examined where A and B represented brief sinusoidal tone bursts, and – represented a silent interval. The A and B bursts differed in frequency and were varied to be closer in frequency to one another, or further in frequency from one another. The effect of instructions on perception were especially examined in the ambiguity region (where listeners were unsure if they heard two separate streams or one integrated stream). When the sound sequence falls between the fission boundary and the temporal coherence boundary (ambiguity region) (Figure A2), listeners report a "flip" spontaneously between integration and segregation (Cooper & Roberts, 2007; Moore & Gockel, 2002). Figure A1 represents the percept of a tone sequence and how it's affected by the frequency separation of the tones (Moore & Gockel, 2012).

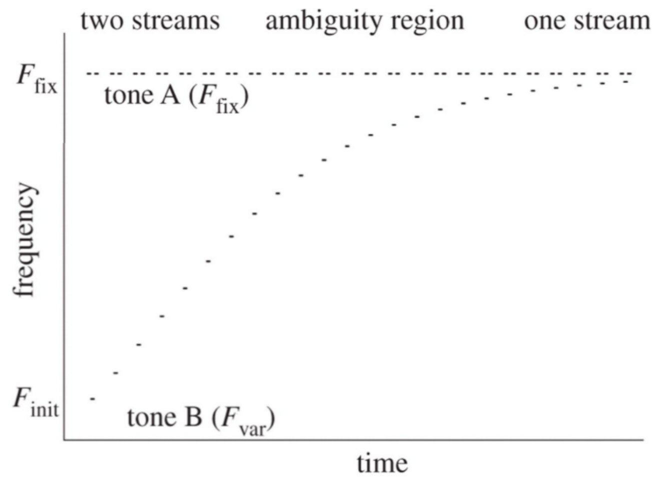

Figure A1. Moore and Gockel (2012) figure of the concept of the ambiguity region.

In the ambiguity region, there was a tendency for fission to occur. Fission occurred with increasing exposure time to the sequences. Moore and Gockel (2002) state, “the auditory system starts with the assumption that there is a single sound source, and fission is only perceived when sufficient evidence has built up to contradict this assumption.” This build-up effect seems to stabilize after around ten seconds. The build-up effect will be discussed further in this section.

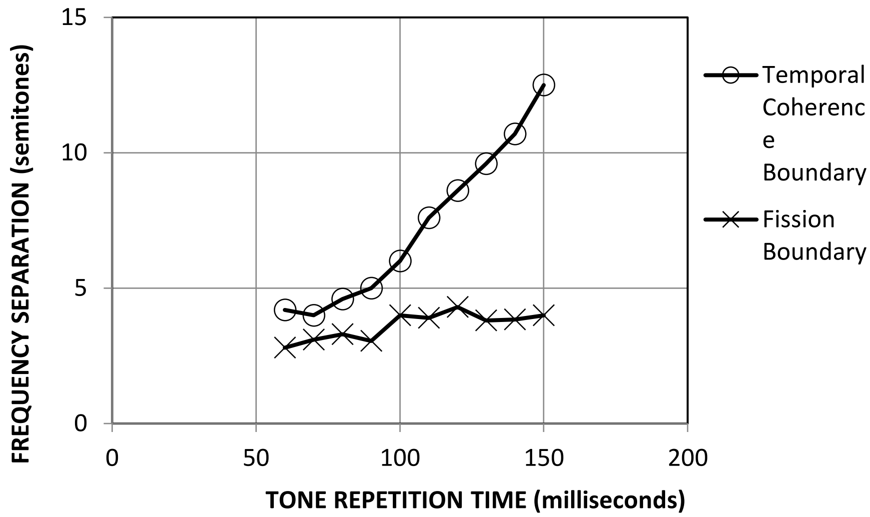

Figure A2. The Fission Boundary (FB) and the Temporal Coherence Boundary (TCB) for auditory stream segregation in van Noorden's unpublished dissertation "Temporal coherence in the perception of tone sequences" (Adapted from van Noorden, 1975).

The perception of two separate streams or one continuous stream is also influenced by task demands. That is, whether segregation abilities are assessed using subjective or objective measures. Experiments assessing stream segregation abilities therefore produce different results when using subjective vs. objective measures. Moore and Gockel (2002) argue for the use of objective measures of stream segregation abilities. The objective approach examines performance using a task that is assumed to be affected by streaming (Moore & Gockel, 2002). Performance is task-dependent -- where performance is expected to be better if fission occurs, or performance decreases if fission occurs (Micheyl & Oxenham, 2010; Moore & Gockel, 2002).

Moore and Gockel (2002) also examined the contribution of the peripheral auditory system in stream segregation. Historically, it has been proposed that streaming depends upon the filtering that takes place at the basilar membrane of the cochlea, primarily. Moore and Gockel (2002) refer to the Beauvois and Medis (1997) computer model, which is based on the idea “that streaming depends upon the overlap of the excitation patterns evoked by successive sounds in the cochlea” (Moore & Gockel, 2002). This model is based on the concept of “peripheral channeling” and has calculated Equivalent Rectangular Bandwidth (ERB) of auditory filters for normal hearing listeners (Moore & Gockel, 2002). This model predicts that “for a rapid sequence of sinusoids, the frequency separation of successive tones at the fission boundary should correspond to a constant difference in ERB number” (Moore & Gockel, 2002). Both frequency separation and pitch influence peripheral channeling, but studies suggest spectral similarity rather than pitch similarity is more important (Moore & Gockel, 2002). It is known that frequency separation of consecutive tones has a strong influence on streaming, which is consistent with the idea that overlap of excitation patterns in the peripheral auditory system plays a strong role (Moore & Gockel, 2002). Van Noorden (1975) also concluded that “contiguity at the level of the cochlear hair cells was necessary for fusion to occur” (Moore & Gockel, 2002). Hartmann and Johnson (1991) found that conditions where successive tones differed in spectrum (i.e. greater peripheral channeling) led to the best performance. Finally, Bregman et al., 2001 investigated the influence of excitation patterns on stream segregation of successive sounds. The tone bursts they presented in this experiment differed in center frequency, but were matched overall in relative bandwidth. Segregation was only slightly affected by bandwidths, even when there was substantial spectral overlap. These results indicated that with increasing differences in center frequency of the two bands came increased perceived segregation. On the other hand, these results also indicated that overlap of excitation patterns does not necessarily prevent segregation.

Overlap of peripheral excitation patterns are not the only factors that contribute to streaming. Rose and Moore (1997) determined this based on their research with hearing impaired listeners. With cochlear hearing losses, comes broader than normal excitation patterns. So, one would assume the fission boundary would be larger for hearing impaired listeners if overlap of peripheral excitation patterns determines the fission boundary (Moore & Gockel, 2002). But, their results indicated that some bilaterally hearing-impaired listeners showed fission boundaries within normal limits. Rose and Moore (1997) discussed the change in the frequency-to-place mapping that results from cochlear hearing loss. Frequency-to-place mapping describes the decrease in the characteristic frequency for a given place with increasing hearing loss (outer hair cell damage). Congruently, as hearing loss increases, the position of peak excitation produced by a specific frequency tone shifts towards the base of the cochlea (Moore & Gockel, 2002). The normal fission boundaries in some hearing-impaired listeners may be explained when the hearing loss is relatively uniform across frequency (all center frequencies shift by a similar ratio). Those hearing-impaired listeners with different fission boundaries than normal may be explained by varying hearing loss across frequency, causing a distortion of the frequency-to-place map (Moore & Gockel, 2002). Other variations in fission boundaries across hearing-impaired listeners may be explained by variation in frequency discrimination abilities.

So what might be the other factors influencing streaming abilities? Moore and Gockel (2002) describe temporal envelope and bandwidth factors, phase factors, fundamental frequency factors, and finally, lateralization factors. Results from studies investigating temporal envelope and bandwidth factors revealed segregation abilities increased when using tone-noise combination sequences vs. noise-noise or tone-tone sequences. This is because tone-noise combinations had a difference in timbre resulted from the difference in the temporal envelope between the noise and the tone (Dannenbring & Bregman, 1976; Moore and Gockel, 2002). Other research supported differences in temporal envelope can enhance stream segregation abilities (Cusack & Roberts, 1999; Vilegen & Oxenham, 1999; Cusack & Roberts, 2000; Moore & Gockel, 2002). Phase factors on stream segregation were also assessed by researchers. Results of these experiments revealed performance was poorer when successive sounds differed in phase spectrum and performance increased when successive sounds had the same phase spectrum (Roberts et al., 2002; Moore & Gockel, 2002). Moore and Gockel (2002) report "the effects of phase are presumably mediated by changes in the waveform or envelope of the sound produced by the phase manipulation." Although phase differences may have an influence on streaming, phase differences do not provide strong cues for stream segregation when a listener is some distance from a sound source (Moore & Gockel, 2002).

Fundamental frequency has been shown to influence streaming abilities. Researchers agreed that listeners use fundamental frequency differences and spectral shape between consecutive tones to achieve stream segregation (when the task is promoting stream segregation) (Bregman et al., 1990; Singh, 1987; Vliegen & Oxenham, 1999, Moore & Gockel, 2002). Other researchers (have found that periodicity differences may either increase or decrease performance on experimental tasks, depending on if the task is promoting segregation or integration. Depending on the task, results can vary in this manner, resulting in conflicting research on this topic (Moore & Gockel, 2002).

Additionally, the difference in amplitude-modulation (AM) rate difference has been shown to be a cue for stream segregation (Grimault et al, 2002, Hong & Turner, 2009, Nie & Nelson, 2015). These studies superimposed AM on either broadband noise (e.g., Grimault et al, 2002) or wideband noise (Nie & Nelson, 2015). Overall, it is suggested that stream segregation can be elicited with the AM-rate difference of 1 octave or larger, but the strength of segregation saturates when the AM-rate difference reaches around 2 octaves. AM has been suggested to be analyzed by the neurons which are tuned to different modulation rates (Kay, 1982) at the central levels (e.g., cochlear nucleus cf. Møller, 1976; inferior colliculus, cf. Rees & Møller, 1983, Lorenzi et al, 1995). Thus, central auditory processing plays a role in the underlying mechanisms of stream segregation. Moreover, streaming abilities may be affected by lateralization factors. Because lateralization depends on more cortical processing, stream segregation as a result of presenting successive stimuli to opposite ears cannot be classified as "peripheral channeling" (Moore & Gockel, 2002). Experiments using interaural time or intensity differences have provided evidence that differences in lateralization influence streaming abilities. Although peripheral channeling has been the most reported factor influencing streaming abilities, these other factors mentioned (i.e., temporal envelope, periodicity, spectral shape, AM-rate, lateralization) have been demonstrated to play a role in streaming abilities. Moreover, it is important to remember that streaming abilities are directly related to the amount of perceptual difference between successive sounds (whether it is frequency, phase, lateralization, etc.). Moore and Gockel (2002) state that "any sufficiently salient perceptual difference may lead to stream segregation, regardless of whether or not it involves peripheral channeling."

Lastly, in addition to the static cues discussed above, the predictability over the course of a stimulus sequence has been shown to play an important role in the formation of auditory streams (for a review, see Bendixen et al., 2014). Temporal rhythmic regularity is of specific interest in the context of the current study. Bendixen et al. (2010) used ABA sequences where individual tones were varied in both frequency and intensity level. These levels were chosen "to preserve similarity within each set while promoting a clear differentiation between the two sets." Random sequences were chosen for each participant, but the regularities were the same for all participants. Participants were instructed to indicate their percept ("segregated," "integrated," "neither," or "both"). Results of this study suggested that "detection of temporal regularities within a given sound organization increases the strength of the dominant organization" (Bendixen et al., 2010). Unlike frequency separation, rhythm does not constitute a primitive cue for stream segregation. Bendixen et al. (2010), state that rhythm is but a secondary factor influencing the stability of a perceptual organization. Finally, Bendixen et al. (2010) report that the rhythmic cue can prolong voluntary stream segregation, while does not appear to affect the integrated percept.

Build-up effect

Auditory stream formation has been shown to be dependent on the amount of time the target sequence is presented. The tendency for segregation (fission) to occur increases with longer exposure time to the sound sequence. Additionally, the fission percept builds up rapidly over about ten seconds, and builds more slowly up to at least 60 seconds (Moore & Gockel, 2012). As mentioned above, the auditory system assumes that a single sound source is available, but with continuous presentation or build-up, the auditory system can separate streams (fission) (Moore & Gockel, 2012). Moore and Gockel (2012) describe several factors as evidenced by research (Anstis & Saida, 1985) toward the presence of the build-up effect. These researchers describe a "resetting" of a fission percept due to build-up, back to a fusion percept when the sound sequence was switched suddenly to the other (contralateral) ear. Rogers & Bregman (1993) researched this "resetting" and supported reduced segregation abilities when there was a shift in perceived location or loudness of the sequence (because of stimuli being presented to the contralateral ear). They repeated a variation of this experiment, and found that fission percepts returned even after the introduction of stimuli to the contralateral ear if the transition from one location to the other was gradual rather than abrupt (Moore & Gockel, 2012). The results of these studies indicate that sudden changes in a sequence causes the activation of a new sound source, causing the percept to "reset" to the initial condition – fusion (Moore & Gockel, 2012).

Another important factor when examining the build-up effect is attention. Although research remains conflicting as to the role of attention, the role should be discussed. Carolyn et al., (2001) investigated this role in a series of experiments. Researchers presented tone sequences (ABA – ABA) to the left ear of listeners for 21 seconds (with no stimulus to the right ear in the baseline condition). The listeners were told to indicate whether they heard a galloping rhythm (integration) or if they heard two separate streams (segregation). There were two experimental conditions; the first condition had two tasks – make judgements of changes in amplitude of the noise bursts presented to the right ear for the first ten seconds, then switch, and make judgments about the tone sequence in the left ear (experimental). The second experiment was a control condition – the condition was still the same two-task condition, but the instructions were different to the listeners. They were told to ignore the noise bursts in the right ear, and only focus on the task and stimuli in the left ear. The results of these studies revealed a similar result for both the baseline and control conditions – the build-up of stream segregation was present. However, the probability of hearing two streams was significantly smaller in the two-task condition (that involved switching tasks), than when the listeners paid attention to the tone sequence in the left ear the entire time. Subsequent studies by Carolyn et al. (2001) found that stream segregation abilities were decreased when a distracting task was introduced. These results indicate evidence for the role of attention in improved performance of build-up stream segregation (Carolyn et al., 2001; Moore & Gockel, 2012). However, Moore & Gockel (2012) point out this attention contribution may be attributed to the idea of "resetting" presented earlier; switching tasks may have actually reset the percept, causing a decrease in segregation abilities. Other researchers have supported this concern from Moore and Gockel (2012) (Cusack et al., 2004; Moore & Gockel, 2012). Although there is some conflicting ideas, researchers still support that build-up of segregation can be decreased by inattention or by a switch of attention (or a combination) (Carolyn et al., 2001; Cusack et al., 2004; Moore & Gockel, 2012).

There have been recent studies that have argued against the presence of the build-up effect. Dieke et al. (2012) challenged the idea that a single sound source is the initial percept for the auditory system. In their experiment, listeners were instructed to indicate their percept as soon as possible and if their percept changed at all during the presented sequence. They revealed that with the largest frequency separations, a two-stream percept was immediately identified by the listeners (probability of 80%), challenging the idea of an initial fusion percept during a sound sequence. Other researchers have reported a similar initial two-stream percept at large frequency separations (Denham et al., 2013). The only increase in two-stream percepts following the initial onset of the sound sequence at the most ambiguous stimulus condition. Dieke et al. (2012) concluded, "a build-up of segregation is not generic and requires ambiguity of the sound sequences to occur" (Bockmann-Barthel et al., 2014).

Micheyl and Oxenham (2010) use their work to offer several explanations for the competing conclusions about the presence of build-up. In a series of three experiments with normal hearing listeners, both subjective and objective data sets were obtained. The first experiment was a subjective measure of stream segregation and perceived judgements were measured under three stimulus parameters (frequency separation, presentation pace, and number of tones) under three separate instruction conditions to evaluate the "attentional set" theory provided by van Noorden (1975). As mentioned above, attention may have an influence on stream segregation abilities in that listeners may have some control over what percept they hear (Micheyl & Oxenham, 2010). The stimuli sequences used in this experiment were similar to that in the other two experiments (objective experiments), in order to make reasonable data comparisons. The temporal position of the tones in all experiments were jittered randomly, resulting in temporally irregular sequences (Micheyl & Oxenham, 2010). In these experiments, listeners were told to wait until the end of the sequence before a judgement was made. When examining results of the build-up effect specifically, the first experiment (subjective) did not show evidence of build-up as the proportion of “two streams” responses did not increase as a function of sequence length. Micheyl and Oxenham (2010) observed that as long as the frequency separation was relatively large, “two streams” responses were observed in the short stimulus sequences as well. Because of the variability amongst studies using subjective methods of assessing stream segregation abilities, Micheyl and Oxenham (2010) followed their subjective experiment with two objective experiments.

The objective experiments were similar, but the first objective experiment (Experiment 2) was designed in a way that stream segregation would hinder performance and the second objective experiment (Experiment 3) was designed that stream segregation improved performance. Methods differed somewhat between these two experiments as in the first objective experiment (Experiment 2) listeners were comparing the timing of the A and B tones and in the second objective experiment (Experiment 3), listeners were instructed to ignore the A tones and just focus on the timing of the B tones. The objective experiments provided measured thresholds of performance to compare to the subjective results in Experiment 1. Results of Experiment 2 were consistent with the idea that (based on the task) stream segregation would hinder performance. These results were consistent with what was seen in Experiment 1. But, results of Experiment 2 revealed that the length of the sequences had no consistent effect on stream segregation

(Micheyl & Oxenham, 2010). In Experiment 3, stream segregation did show to improve performance (measured thresholds decreased as frequency separations increased) and were better for the faster rate sequences (consistent with previous evidence). However, results of this experiment revealed that performance did not improve as the sequence length increased. This result was consistent with the subjective experiment (Experiment 1), which also showed no significant effect of sequence length on stream segregation. In fact, even when listeners were instructed to actively try to segregate streams, stream segregation still occurred almost immediately after the start of the stimulus sequence (Micheyl & Oxenham, 2010). It appears that under these conditions, evidence toward the build-up effect was not observed.

Although this data in normal hearing listeners is contradictory to previous listeners, Micheyl & Oxenham (2010) offer several reasons for their results. In this study, researchers used tones that were temporally jittered, most previous studies evaluating the build-up effect did not (the stimuli were temporally regular). Micheyl and Oxenham (2010) explain that this jittering may have altered the build-up as evidenced by Okada and Kashino (2008) who found that auditory streaming was reduced when temporal jittering was used. But, conflicting evidence has also been discussed – temporal jittering has been found to have no significant effect on single-stream percepts; therefore, it is unclear as to if temporally irregular sequences has an effect on build-up (Roberts et al., 2008; Micheyl & Oxenham, 2010). Micheyl and Oxenham (2010) propose other explanations for their conflicting results. In the current experiments, relatively short stimulus sequences were used (a few seconds or less compared with ten seconds or more in other studies). Build-up could have been affected in different ways. First, listeners may have used the strategy of paying close attention at the beginning of each sequence to form judgments rapidly (there was an “enhanced preparedness”), which would have by-passed or accelerated the build-up. Next, this study had the listeners indicate their percept after the end of the stimuli sequences, while other studies had listeners indicate their percept during the course of a sequence. Because of this, listeners may have had the opportunity to think about the stimulus they had just heard before responding – this additional time to decide or “mental rehearsal” could have influenced the decision outcome (Micheyl & Oxenham, 2010).

Overall, evidence toward the existence of the build-up effect is still relatively unclear. Because of the variation in methodologies used amongst studies, consistent results have not been found. The manner in which data is collected, has had an influence on results; that is, using subjective vs. objective measures to assess build-up of stream segregation. Additionally, how the stimulus sequences are constructed play a role in obtained results. To show clear evidence either for or against the presence of the build-up effect, studies must use similar objective measures and stimuli constructs in order to draw accurate conclusions (and to be able to make comparisons between studies). Even less clear are the abilities of cochlear implant users to segregate auditory streams. Studies have begun to explore this topic in cochlear implant users, especially when evaluating the build-up effect.

Stream segregation in Cochlear Implant (CI) users

Per the theory of peripheral channeling, stream segregation abilities depend on the amount of overlap in the excitation pattern on the basilar membrane induced by the overlaid sound sequences – the more the sound sequences overlap, the more likely for fusion to occur (Marozeau et al., 2013). Because hearing loss can increase the region of excitation along the basilar membrane due to the impairment of basilar membrane mechanisms, it is inferred that hearing-impaired listeners should show a decrease in stream segregation abilities (Marozeau et al., 2013). Likewise, current spread by a cochlear implant causes a wide stimulation area around each electrode. Because of this, reduced streaming abilities are hypothesized in cochlear implant users (like hearing-impaired users) (Marozeau et al., 2013).

Cochlear implant (CI) users have the ability to understand speech quite well in a quiet environment, but often have a more difficult time in background noise and with music perception. A major limitation of cochlear implant benefit is speech understanding – speech understanding is greatly reduced by competing background sounds. Nelson and Jin (2004) reported competing speech can decrease performance in implant users even at a +16 dB signal-to-noise ratio (Cooper & Roberts, 2007). This can be attributed to the varying spatial interaction between electrode channels (Cooper & Roberts, 2007). The spread of excitation across channels can lead to a high degree of spectral "smearing" and reduced spectral resolution (Cooper & Roberts, 2007). Timbre discrimination and pitch discrimination are two common degradations related to cochlear implant use. Perceptual differences between sound sources are reduced with the use of a cochlear implant and therefore result in reduced auditory streaming abilities. Because of this, not only is understanding speech in a noisy environment difficult, but so is music perception.

Music perception is degraded with cochlear implant use because of the reduction in the ability to distinctly hear multiple and separate lines of a melody, as well as different instruments (Marozeau et al., 2013). Marozeau et al. (2013) researched the effects of a cochlear implant on different acoustic and perceptual cues responsible for streaming to determine a relationship between these differences in cues and melody segregation. In the first experiment, the listeners were told to direct their attention toward segregation ("finding" the melody). This experiment generated a subjective measure of streaming. The sound sequences consisted of a fixed melody with overlaid "distractor" notes which varied in intensity, fundamental frequency range, temporal envelope, or spectral envelope. Results overall of this experiment indicated that, "as the physical different between the target and the distractor increased, listeners on average reported less difficulty segregating the melody from distractor notes" (Marozeau et al., 2013). This finding is relatively consistent with previous research in normal-hearing listeners (described above). The marked difference in these results compared to normal-hearing listeners was in the intensity parameter -- normal hearing listeners required more attenuation of the distractor notes in order to reach the same perceived difficulty level as the cochlear implant users. Marozeau et al. (2013) explain this result could be because of the steep loudness-growth function with electrical stimulation experienced by cochlear implant users – "they experience more change in loudness for a given physical change than in normal hearing" (Marozeau et al., 2013). Other interesting results of this experiment included that although CI listeners needed a larger difference in spectral envelope to perceive the target melodies than normal-hearing listeners, CI users reported similar difficulties as normal-hearing listeners in the task when fundamental frequency was the parameter of interest. Overall, results from this study suggest that like normal-hearing listeners, cochlear implant users can segregate streams when the difference between them is large enough. Additionally, although streaming abilities were different between the groups, CI users were in fact able to segregate melodies using perceptual differences generated by each acoustic cue (Marozeau et al., 2013).

Other subjective studies of stream segregation abilities in cochlear implant listeners have been conducted. Chatterjee et al. (2006) report similar results; from preliminary studies, CI listeners are able to use intensity differences to segregate stimuli sequences. They used stimuli sequences that were varied (in order to measure build-up) and consisted of two pulse trains which were different in cochlear location. As the frequencies of the sequences became further apart, streaming became stronger (consistent to above results). However, although they found evidence toward streaming abilities in cochlear implant users, no build-up effect was evidenced. Chattergee et al. (2006) concluded that “it is possible for the electrically stimulated auditory system to perceptually segregate stimuli based on differences in either cochlear place or temporal envelope” (Chatterjee et al., 2006).

Additionally, Bockmann-Barthel et al. (2014) compared cochlear implant users to normal-hearing listeners when evaluating the build-up of streaming. They evaluated the proportion of time for distinguishing the perception of one-stream vs. two-streams and the time needed to make that first perceptual decision. Stimuli was presented in the soundfield and used ABAB sequences with different frequency separations. Listeners were instructed to select one-stream or two-streams as soon as they had a percept and to switch whenever that percept changed. Bockmann-Barthel et al. (2014) concluded similar results between cochlear implant users and normal-hearing listeners – as frequency separations increased, a two-stream percept increased. When evaluating the build-up effect, Bockmann-Barthel et al. (2014) reported the traditional hypothesis of build-up (a two-stream percept emerges from an initial one-stream percept over time), was not supported. That is, users did not default to a one-stream percept at first, then later change their response to a two-stream percept. In fact, in this study, both normal-hearing listeners and cochlear-implant listeners defaulted to favor the two-stream percept initially (Bockmann-Barthel et al., 2014).

Although several studies have provided evidence of stream segregation abilities in cochlear implant users, others have argued that cochlear implant users do not show indication of stream segregation abilities. Cooper and Roberts (2007) reason that cochlear implant listeners must rely on schema-based processing to separate a perceptual stream which puts them at a considerable disadvantage in complex listening environments (especially if attentional resources are limited). This is a contrast to Chatterjee et al. (2006), who suggested that cochlear implant users may have primitive stream segregation abilities. Cooper and Roberts (2007) conducted a study to examine streaming abilities in cochlear implant listeners. They examined the effect of frequency separation, the effect of rate of presentation on the probability of stream segregation, and the ambiguous perceptual region (described above). Again, this experiment involved subjective responses to the perceptions of segregated vs. integrated stimuli sequences. Cooper and Roberts (2007) concluded that although cochlear implants provide listeners the ability to discriminate between subsets of sounds in a sequence, automatic segregation does not necessarily occur. Because frequency separation was the only parameter to show improved segregation abilities, Cooper and Roberts (2007) mention that it is not possible to conclude that is a genuine measure of stream segregation without a significant effect of the other two parameters (as all three are properties found in normal-hearing listeners who show automatic stream segregation). Cooper and Roberts performed a follow-up study in 2009 and reported similar conclusions to their study in 2007 – cochlear implant users do not provide evidence of automatic stream segregation and again suggest that cochlear implant users may “instead rely on a schema-based selection to hear a subset of acoustic elements from a sequence as a separate stream” (Cooper & Roberts, 2009).

This current study is a follow-up study to Nie and Nelson (2015). Nie and Nelson (2015) objectively evaluated the role of spectral overlap and amplitude-modulation rate on build-up stream segregation in normal-hearing listeners where the stimuli was constructed to resemble the spectral interaction of signals delivered through a cochlear implant. Nie and Nelson (2015) used long (12-pair) and short (3-pair) sequences, where the long sequences (12-pair) were used to evaluate for the build-up effect. A and B bursts were either bandpass or broadband noise carrying sinusoidal amplitude modulation and differed in either the center frequency of the noise band, in amplitude-modulation rate, or in both. In the overlaid sequences, A bursts were jittered from their nominal temporal locations by a random amount (ranging from 0 to 40 ms) and B bursts were held steady at a lower frequency range. The last B burst in the sequence was either delayed from its nominal temporal location by 30 ms, or it was not delayed (or advanced by a random amount ranging from 0 to 10 ms). Listeners were presented the stimuli through TDH-49 headphones monaurally to the right ear at 70 dB SPL and the listening task included identifying if a delay or a no-delay sequence was presented; feedback was given to the listeners directly after each sequence. Performance (d’) was measured through a single-interval yes/no approach. This task forced the listeners to direct their attention on segregating streams in order to correctly select the delay vs. no-delay sequence. Because the A bursts were jittered, an uncertainty was introduced to the listeners; therefore, the A-to-B gap was an ineffective cue for identifying delayed B bursts. The better the listeners could segregate the bursts, the better their performance in identifying delay vs. no-delay bursts. Results of this study indicated that spectral separation for wide-bandpass noises induced stream segregation. Additionally, amplitude-modulation rate differences aided in stream segregation. Amplitude-modulation rate differences were not found to have as large of an effect on streaming abilities as spectral separations. Furthermore, build-up of stream segregation was present with adequate spectral separation and amplitude-modulation rate differences. This means that higher d' values were found for 12-pair (long) sequences than for the 3-pair (short) sequences. These results suggest that CI users may show stream segregation abilities if spectral separations and amplitude-modulation rate differences are large enough (Nie & Nelson, 2015). Based on these results, the current study was created with similar stimuli constructs (with modifications) and procedures to evaluate the build-up of streaming in cochlear implant users.

References

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulation tones. *F. Exp. Psychol.: Human Percept. Perf., 11*, 257-271. doi: 10.1037/0096-1523.11.3.257

Beauvois, M.W. & Meddis, R. (1997). Time decay of auditory stream biasing. *Percept. Psychophsy., 60(1998)*, 1216-1227.

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. [Review]. *Front. Neurosci., 8.* doi: 10.3389/fnins.2014.00060

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. The Journal of the Acoustical Society of America, 128(6), 3658-3666.Bockmann-Barthel, M., Deike, S., Brechmann, A., Ziese, M., & Verhey, J.L. (2014). Time course of auditory streaming: do CI users differ from normal-hearing listeners. *Front. Psychol., 5(775)*, 1-9. doi: 10.3389/fpsyg.2014.00775

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Bregman, A.S., Ahad, P.A., Van Loon, C. (2001). Stream segregation of narrow-band noise bursts. *Percept. Psychophys., 63,* 790-797

Bregman, A. S., Colantonio, C., & Ahad, P. A. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Percept Psychophys, 61(2)*, 195-205. doi: 10.3758/BF03206882

Bregman, A.S., Liao, C., Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canad. J. Psyschol., 44,* 400-413

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *J. Acoust. Soc. Am., 60,* 863–869.

Burns, E. M., & Viemeister, N. F. (1981). Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *J. Acoust. Soc. Am., 70*, 1655–1659.

Carolyn, R.P., Cusack, R., Foxton, J.M. & Robertson, I.H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *F. Exp. Psychol.: Human Percept. Perf., 27,* 115-127. doi: 10.1037/0096-1523.27.1.115

Chatterjee, M., Sarampalis, A. ,& Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. *Hear. Res. 222*, 100–107. doi: 10.1016/j.heares.2006.09.001

Cooper, H.R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. *Hear. Res., 225(2007),* 11-24. doi: 10.1016/j.heares.2006.11.010

Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception, 28,* 1281-1289

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential group. *Percept. Psychophys., 62,* 1112-1120

Cusack, R., Deeks, J., Aikman, G. & Carolyn, R.P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *F. Exp. Psychol.: Human Percept. Perf., 30,* 643-656. doi: 10.1037/0096-1523.30.4.643

Dannenbring, G.L., & Bregman, A.S. (1976). Stream segregation and the illusion of overlap. *J. Exp. Psychol.: Human Percept. Perf., 2*, 544-555, doi: 10.1037/0096-1523.2.4.544

Denham S. L., Gyimesi K., Stefanics G., Winkler I. (2013). Perceptual bistability in auditory streaming: how much do stimulus features matter? *Learn. Percept. 5,* 73–100. doi: 10.1556/LP.5.2013.Suppl2.6

Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *J. Acoust. Soc. Am., 128(1),* EL1-EL7. doi: doi:http://dx.doi.org/10.1121/1.3436498

Dieke, S., Heil, P., Bockmann-Barthel, M., & Brechmann, A. (2012). The build-up of auditory stream segregation: A different perspective. *Front. Psychol., 3(461),* 1-7. doi: 10.3389/fpsyg.2012.0046

Fitzgerald, M. B., & Wright, B. A. (2005). A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *J. Acoust. Soc. Am., 118(6),* 3794-3803. doi: 10.1121/1.2074687

Hartmann, W.M. & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Percept., 9,* 155-184. doi: 10.2307/40285527

Jesteadt, W. (1980). An adaptive procedure for subjective judgments. *Percept. Psychophys., 28,* 85-88. http://dx.doi.org/10.3758/BF03204321

Kay, R. H. (1982). Hearing of modulation in sounds, *Physiol. Rev*. 62, 894-975.

Lemańska, J., Sęk, A., & Skrodzka, E. (2002). Discrimination of the amplitude modulation rate. *Arch. Acoust., 27(1),* 3-22.

Li, X., Liang, Z., Kleiner, M., & Lu, Z. L. (2010). RTbox: A device for highly accurate response time measurements. *Behav Res Methods, 42(1),* 212-225.

Lorenzi, C., Micheyl, C., and Berthommier, F. (1995). Neuronal correlates of perceptual amplitude-modulation detection, *Hear. Res., 90,* 219-227.

Macmillan, N.A., and Creelman, C.D. (2005). Detection Theory: A User's Guide. Mahwah, NJ: Lawrence Erlbaum Associates

Marozeau, J., Innes-Brown, H., & Blamey, P.J. (2013). The acoustic and perceptual cues affecting melody segregation for listeners with a cochlear implant. *Front. Psychol., 4(790),* 1-11. doi: 10.3389/fpsyg.2013.00790

Micheyl, C., & Oxenham, A.J. (2010). Objective and subjective psycholphysical measures of auditory stream integration and segregation. *J. Assoc. Res. Otolaryngol., 11,* 709-724. doi: 10.1007/s10162-010-0227-2

Møller, A. R. (1976). Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus, Act Physiologica Scandinavica. 98, 157-167.

Moore, B.C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Ac., 88(2002),* 320-332.

Moore, B.C., & Gockel, H. (2012). Properties of auditory stream formation. *Phil.Trans. R. Soc. B, 367*, 919-931. doi: 10.1098/rstb.2011.0355

Nie, Y., & Nelson, P.B. (2015). Audiotry stream segregation using amplitude modulated bandpass noise. *Front. Psychol., 6(1151),* 1-10. doi: 10.3389/fpsyg.2015.01151

Nie, Y., Zhang, Y., & Nelson, P. B. (2014). Auditory stream segregation using bandpass noises: evidence from event-related potentials. [Original Research]. *Front. Neurosci., 8.* doi: http://journal.frontiersin.org/Journal/10.3389/fnins.2014.00277/full

Okada M., & Kashino M. (2008). Temporal jitter affects perceptual transitions in auditory streaming. Proceedings of the 31st MidWinter Research Meeting of the Association for Research in Otolaryngology

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision*, *10*, 437-442.

Rees, A. and Møller, A. R. (193). Responses of neurons in the inferior colliculus of the rat to AM and FM tones, *Hear. Res., 10*, 301-310.

Roberts, B., Glasberg, B.R., & Moore, B.C.J. (2002) Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am., 112,* 2074. doi: 10.1121/1.1508784

Roberts B., Glasberg B.R., Moore B.C. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *J. Exp. Psychol.: Hum. Percept. Perform., 34*, 992–1006

Roger, W.L., & Bregman, A.S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Percept. Psychophys., 53,* 179-189. doi: 10.3758/BF03211728

Rose, M.M, & Moore, B.C.J. (1997). Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *J. Acoust. Soc. Am., 102*, 1768-1778Singh, P.G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre. *J. Acoust. Soc. Am., 82*, 886-895

Van Noorden, L.P.A.S. (1975). Temporal coherence in the perception of tone sequences. Unpublished Ph.D. dissertation, Eindhoven University of Technology.

Vliegen, J., & Oxenham, A.J. (1999). Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am, 105,* 339-346

Walker, G., Dillon, H., & Byrne, D. (1984). Sound field audiometry: recommended stimuli and procedures. *Ear Hear, 5(1)*, 13-21.